INTRODUCTION

When I am asked what I do for a living and I say that I test spam filters, people often start telling me how much they dislike spam and how they can’t imagine why people fall for Nigerian 419 scams. Next comes the question: ‘So, what spam filter do you recommend?’

I always find myself hesitating to answer that question. Not just because the tests I run don’t focus on end-user solutions, and not because I wouldn’t be able to name the products that have performed well many tests in a row – I would. But, while that’s important, it isn’t all that matters.

For some organizations it is essential that the product is hosted on their premises. For others it is important that the servers that store their email are located in an EU country. To some it will matter most that as much malware as possible is blocked before the email reaches their systems, while others will find it essential that the product can be controlled using some kind of API.

We have always stressed that a decent spam catch rate and false positive rate are not the be-all and end-all of a good spam filter. They are, of course, very important and will continue to be the main focus of the VBSpam tests but, starting with this test, we will also comment on a number of qualitative features of participating products – such as whether anti-malware scanning is included and, for hosted solutions, whether multiple MX records are supported. This will give potential customers more information about the products.

This month also sees the introduction of a new award, ‘VBSpam+’. This will be awarded to those products that block all but at most one in 200 spam messages (i.e., have a spam catch rate of 99.5% or higher) and at the same time do not block any legitimate email. In this test one product earned a VBSpam+ award.

All of the other participating full solutions achieved a VBSpam award, but most of them saw a deterioration in performance compared to the previous test. The test also included two partial solutions (DNS blacklists), designed to be used in conjunction with other products to provide anti-spam protection.

THE TEST SET-UP

The VBSpam test methodology can be found at http://www.virusbtn.com/vbspam/methodology/. As usual, email was sent to the products in parallel and in real time, and products were given the option to block email pre-DATA. Five products chose to make use of this option.

As in previous tests, the products that needed to be installed on a server were installed on a Dell PowerEdge R200, with a 3.0GHz dual core processor and 4GB of RAM. The Linux products ran on SuSE Linux Enterprise Server 11; the Windows Server products ran on either the 2003 or the 2008 version, depending on which was recommended by the vendor.

To compare the products, we calculate a ‘final score’, which is defined as the spam catch (SC) rate minus five times the false positive (FP) rate. Products earn VBSpam certification if this value is at least 97:

\[ SC - (5 \times FP) \geq 97 \]

As mentioned, products can earn VBSpam+ certification if they combine a spam catch rate of 99.5% or higher with a zero false positive rate.

THE EMAIL CORPUS

The test ran for 16 consecutive days, from 12am GMT on Saturday 11 February 2012 until 12am GMT on Friday 27 February 2012.
The corpus contained 201,049 emails, 191,893 of which were spam. Of these, 111,296 were provided by Project Honey Pot and 80,597 were provided by Spamfeed.me, a product from Abusix. They were all relayed in real time, as were the 8,952 legitimate emails (‘ham’) and the remaining 204 emails, which were all legitimate newsletters.

Looking at the spam corpus we noticed that emails linking to websites selling (fake) Viagra and other pharmaceutical products continue to make up a significant part of it. Interestingly, we noticed spammers using templates for many social networking sites (Facebook, Twitter, MySpace and Habbo) as well as Amazon to link to such sites, which raises the question of whether rogue pharmacies are more profitable for cybercriminals than social network credentials.

Figure 1 shows the catch rate of all full solutions throughout the test. To avoid the average being skewed by poorly performing products, the highest and lowest catch rates have been excluded for each hour.

In the previous test, the hourly average rarely dipped below a 99.5% catch rate, but in this test it happened regularly, especially during the first week, when average catch rates dropped well below 99% twice. On 13 February this was mostly due to a bunch of German language Viagra spam. Four days later, it was German spam once again that caused a problem for several products; this time the messages were phishing for credit card details.

We didn’t just see a few dips in spam filters’ performance: for all but one product the catch rates were lower than in the previous test. Looking at the various pre-DATA scores (which were also much lower than in previous tests), we see that products had a significantly harder time blocking messages based on the IP address from which they were sent. This may be because DNSBLs have become less accurate than they used to be, or because spammers have been sending more spam using legitimate mail servers (whose IP addresses thus cannot be blacklisted).

This test also saw a significant increase in the size of the ham corpus, which consisted of emails sent to well over 100 email discussion lists, on a broad range of topics and in various languages. Many new sources were added close to the start of the test, which meant that product developers did not receive any feedback on their performance. Helping developers improve their products is an important part of the VBSpam tests, and feedback is essential for that, but we

Figure 1: Spam catch rate of all full solutions throughout the test period.
wanted to test the products’ ability to distinguish legitimate emails from spam without having any prior knowledge of them.

MALWARE, INTERFACES AND LOCATIONS

This month, the MIME protocol celebrates its 20th birthday. It is hard to think of email being as successful as it has been without MIME, which among other things allows for emails to be sent in different character sets and for files to be attached to emails.

The latter property means that, in principle, anyone can send any file to any Internet user, and this has been heavily abused by malware authors to spread their creations. Recently, it seemed that email attachments had stopped being a major method for malware to spread – until last autumn, when malicious spam campaigns suddenly returned: the attachments are usually zip files containing executables masquerading as PDFs or Word documents.

It isn’t just large campaigns that should worry email users. Email is commonly used as the first infection vector in targeted attacks: it is not hard to imagine how an attachment with what looks like relevant information received from what appears to be a known sender could be opened without giving it much thought. Indeed, this is how many well known targeted attacks started, from the RSA hack to ‘Duqu’.

Thus the suggestion that we test products’ ability to block malicious attachments is an understandable one. However, anti-malware testing is not a trivial thing to do, and definitely not something that should be done ‘on the side’.

Instead, we have listed whether products scan incoming emails for malware and, if they do, which engine(s) are used. For those wanting to know how well these solutions detect malware, we refer readers to our VB100 tests, where the on-demand results are of most relevance for the scanning of email attachments.

It is important to note here that many malicious emails are blocked by spam filters without the need for anti-malware scanning, and also that if a malicious attachment does make it into the user’s inbox, its computer is not infected until the attachment is opened¹ (and even then it may be picked up by anti-malware software running on the machine). Thus, while we stress the importance of running anti-malware solutions, we do not wish to make a statement about its necessity in spam filters.

We didn’t just look at anti-malware scanning. For ‘local solutions’ (that is, those located on the user’s premises), we also looked at how an administrator can manage the solution: via a command-line interface (CLI); via a graphical user interface (GUI) running as a desktop program; via a web interface (web GUI); or via an API. We don’t want to make judgements about what kind of interface is better: this depends both on the user’s preferences and an organization’s requirements. Of course, many products can be controlled via more than one interface.

For hosted solutions – all of which can be managed via a web interface – we looked at something else: whether the product supports multiple MX records and whether the product’s mail servers are located at different physical locations. Both could ensure that email arrives even when one or more servers are down or for some reason unreachable. For organizations for which email is critical to the running of the business, this could be an important property.

Interestingly, in speaking to several participants, we learned that the physical location can be important for other reasons as well: data stored on a server is subject to local legislation and for this reason some organizations want to avoid using servers located in certain countries, or only use servers located in specific countries. We did not list where the products’ servers are located; potential customers to whom this is important are advised to contact the vendors themselves.

RESULTS

In the text that follows, unless otherwise specified, ‘ham’ or ‘legitimate email’ refers to email in the ham corpus – which excludes the newsletters – and a ‘false positive’ is a message in that corpus that has been erroneously marked by a product as spam.

The ‘false negative rate’, mentioned several times in the text that follows, is the complement of the spam catch rate: the percentage of spam messages that were not blocked. It can be computed by subtracting the SC rate from 100% – the lower the false negative rate, the better.

Because the size of the newsletter corpus is significantly smaller than that of the ham corpus, a missed newsletter will have a much greater effect on the newsletter false positive rate than a missed legitimate email will have on the false positive rate for the ham corpus (e.g. one missed email in the ham corpus results in an FP rate of 0.02%, while one missed email in the newsletter corpus results in an FP rate of 0.5%).

¹ Occasionally, we hear scare stories about malicious email attachments being able to infect users without being opened. While it is not unimaginable for a vulnerability in an email client to allow this to happen, in most cases the stories have turned out to be false.
AnubisNetworks Mail Protection Service

**SC rate:** 99.92%
**FP rate:** 0.04%
**Final score:** 99.70

*Project Honey Pot SC rate:* 99.9%
*Abusix SC rate:* 99.96%

*Newsletters FP rate:* 0.0%

As with most hosted solutions, AnubisNetworks offers customers multiple MX records spread over multiple locations. File attachments are scanned with ClamAV.

However, it’s likely that few malicious emails will need to be scanned as the product’s spam catch rate has traditionally been high. This month it is slightly lower than in the previous test, but at 99.92% this is hardly an issue. With four false positives, it performs better than average too and thus a tenth VBSpam award goes to AnubisNetworks with the fourth highest final score.

---

BitDefender Security for Mail Servers 3.0.2

**SC rate:** 98.94%
**FP rate:** 0.00%
**Final score:** 98.94

*Project Honey Pot SC rate:* 99.45%
*Abusix SC rate:* 98.23%

*Newsletters FP rate:* 0.5%
Unsurprisingly, BitDefender’s anti-spam solution uses the BitDefender anti-malware engine to check for malicious attachments. The product has been running non-stop in our test lab for about three years, so it uses an older interface which can only be controlled via the command line. However, the most recent version of the product can be controlled via a web GUI as well. (The engine of the product included in the test is, of course, the most recent version.) Like most products this month, BitDefender saw its spam catch rate drop significantly to just below 99%. However, the product was one of only three full solutions to have a zero false positive rate, so its VBSpam award was assured and the product is the only one to have won a VBSpam award in all 18 tests.

Fortinet FortiMail

SC rate: 99.45%
FP rate: 0.04%
Final score: 99.23
Project Honey Pot SC rate: 99.44%
Abusix SC rate: 99.48%
Newsletters FP rate: 0.5%

Fortinet’s FortiMail appliance can be managed via a web GUI, but those who prefer to stay on the command line will be happy to learn that this is also supported via SSH and telnet. Attachments are scanned using Fortinet’s own anti-virus engine and signature database.

As was the case with many products in this test, FortiMail’s false negative rate doubled from its last score. The product also blocked four legitimate emails, but these were the first false positives from the product in six months, so it is possible that this was a temporary glitch. Fortinet earns its 17th VBSpam award.

GFI MailEssentials

SC rate: 99.75%
FP rate: 0.04%
Final score: 99.53
Project Honey Pot SC rate: 99.72%
Abusix SC rate: 99.79%
Newsletters FP rate: 0.0%

GFI’s MailEssentials product currently does not scan attachments for malware, but a newer version of the product – to be released this spring – will do so using the company’s MailSecurity engine. Administrators can manage the current product via a GUI – that is, a snap-in to the Microsoft Management Console. Compared to the previous test, GFI’s false negative rate doubled, but at 99.75% its spam catch rate is still higher than average. A reduced false positive rate saw the product achieve a good final score and GFI earns another VBSpam award.

Halon Security

SC rate: 99.43%
FP rate: 0.03%
Final score: 99.26
Project Honey Pot SC rate: 99.47%
Abusix SC rate: 99.37%
Newsletters FP rate: 0.5%

We have previously praised the way Halon gives administrators flexibility by allowing them to actually program the product via its own scripting language. It is not surprising that this scripting language offers an API via SOAP. Of course, the program also offers a web GUI and a command-line interface for those who prefer the old-fashioned way of managing a product. Attachments are scanned with three anti-malware engines: Kaspersky, Commtouch RPD and ClamAV.

Although we saw a drop in spam catch rate, the drop was smaller than for most other products and the product’s false positive rate dropped as well – resulting in an improvement in its final score. Halon’s seventh consecutive VBSpam award is thus well deserved.

IBM Lotus Protector for Mail Security

SC rate: 99.72%
FP rate: 0.08%
Final score: 99.33
Project Honey Pot SC rate: 99.7%
Abusix SC rate: 99.75%
Newsletters FP rate: 0.0%

As with most appliances, IBM’s virtual appliance can be managed using a web GUI, while the more ‘hardcore’ administrators will be glad to learn that it can also be controlled via the command line. Email attachments are scanned using the Sophos anti-malware engine. IBM’s false negative rate doubled, but with only one in 360 spam messages being missed, the spam catch rate was still rather good. Seven false positives was about average in this test and the product adds its fourth consecutive VBSpam award to its tally.
Kaspersky Anti-Spam 3.0

**SC rate:** 99.57%
**FP rate:** 0.02%
**Final score:** 99.46

**Project Honey Pot SC rate:** 99.54%
**Abusix SC rate:** 99.61%
**Newsletters FP rate:** 0.0%

As a Linux product – the version we use runs alongside the Postfix MTA – it will come as little surprise to learn that Kaspersky Anti-Spam can be managed using a command-line interface. The version of the product we ran does not scan attachments for malware.

As usual, the product had one of the lowest false positive rates and although there was a drop in its spam catch rate, Kaspersky Anti-Spam easily achieved its 16th VBSpam award.
Libra Esva 2.6
SC rate: 99.97%
FP rate: 0.00%
Final score: 99.97
Project Honey Pot SC rate: 99.96%
Abusix SC rate: 99.98%
SC rate pre-DATA: 96.35%
Newsletters FP rate: 0.0%

Version 2.6 of Libra Esva’s virtual anti-spam product was provided for this test. It can be managed using a web GUI, while attachments are scanned for malware using ClamAV.

After achieving the highest final score in the previous test, it seemed like there was little opportunity for Libra Esva to improve. Yet the product did just that: it was the only product in this test to increase its spam catch rate (to 99.97%), while once again blocking none of the now almost 9,000 legitimate emails. The product achieved the highest final score and is the first winner of a VBSpam+ award.

McAfee SaaS Email Protection
SC rate: 99.93%
FP rate: 0.21%
Final score: 98.87
Project Honey Pot SC rate: 99.89%
Abusix SC rate: 99.99%
Newsletters FP rate: 3.9%

McAfee’s anti-malware engine is also used by the company’s hosted solution, which offers customers multiple MX records which can point to different data centres if required.

A small drop in spam catch rate wasn’t too much of a problem as it continued to be over 99.9%. The false positive rate was more of a concern, however, as 19 legitimate emails were blocked – more than any other product. While the product achieved a high enough final score to earn another VBSpam award, we hope that the next test will show that this was a temporary glitch.

Messaging Architects M+Guardian
SC rate: 99.94%
FP rate: 0.18%
Final score: 99.05
Project Honey Pot SC rate: 99.92%
Abusix SC rate: 99.97%
SC rate pre-DATA: 95.22%
Newsletters FP rate: 13.7%

As is common for (virtual) appliances, M+Guardian can be managed using a web interface. Malware is also scanned for attachments, using proprietary tools.

In the last test M+Guardian had a rather high false positive rate. It was nice to see this drop quite a bit – although there is still room for a further reduction. The product’s spam catch rate continues to be very high and wasn’t much lower than in the previous test. Together they pushed the final score to over 99, earning Messaging Architects another VBSpam award for its appliance.

OnlyMyEmail’s Corporate MX-Defender
SC rate: 99.99%
FP rate: 0.02%
Final score: 99.88
Project Honey Pot SC rate: 99.99%
Abusix SC rate: 99.99%
Newsletters FP rate: 2.0%
Given *OnlyMyEmail*'s impressive record of blocking all but a handful of spam messages, one might be tempted to think that malware scanning is unnecessary for this product, yet it was nice to see it offered to those customers who want it, using a mix of proprietary technologies. The hosted solution also offers multiple MX records and the company can balance between locations dynamically to spread the load (this is not something customers control).

The increase in false negatives for *OnlyMyEmail* – from four in the previous test to 12 – is unlikely to be statistically relevant and with a spam catch rate of slightly higher than 99.99% it yet again outperformed all other products on this metric. The two false positives prevented it from achieving a VBSpam+ award, but the FP rate was still far lower than average and earned *OnlyMyEmail* its ninth consecutive VBSpam award with the second highest final score.

**Sophos Email Appliance**

| SC rate: 99.57% | FP rate: 0.07% | Final score: 99.24 |
| Project Honey Pot SC rate: 99.45% | Abusix SC rate: 99.73% | Newsletters FP rate: 0.0% |

As *Sophos* started as an anti-virus company more than a quarter of a century ago, it comes as no surprise that its own anti-malware engine is used in its email appliance. A web interface can be used to control the product.

As with most other products in this test *Sophos Email Appliance* saw its false negative rate increase – doubling from the previous test’s score. At the same time, however, its false positive rate was halved, resulting in an improved final score and the company’s 13th VBSpam award in as many tests.

**SPAMfighter Mail Gateway**

| SC rate: 99.28% | FP rate: 0.07% | Final score: 98.95 |
| Project Honey Pot SC rate: 98.90% | Abusix SC rate: 99.82% | Newsletters FP rate: 3.9% |

Readers of the VB100 reviews will have seen *SPAMfighter*’s sister product *VIRUSfighter* in the reports; it is this product that can be added as an optional extra by those who want their email attachments scanned for malware. The product can be managed using a web GUI.

Compared to the November 2011 review (*SPAMfighter* skipped the January test), the product’s spam catch rate was quite a bit lower, but the decrease in its false positive rate was much more remarkable. With a nicely improved final score, *SPAMfighter* earns its 14th VBSpam award.

**SpamTitan**

| SC rate: 99.93% | FP rate: 0.04% | Final score: 99.71 |
| Project Honey Pot SC rate: 99.92% | Abusix SC rate: 99.95% | Newsletters FP rate: 0.5% |

*SpamTitan*’s virtual anti-spam appliance can be controlled via a web browser or via the command line. Moreover, there is an API that should give administrators even more options.
Attachments are scanned for malware using both Kaspersky and ClamAV.

With a spam catch rate of 99.93% it is barely worth mentioning that SpamTitan’s false negative rate actually doubled compared to the previous test. With just four false positives, and the third highest final score, the company earns its 15th VBSpam award.

Spider saw both its false negative rate and its false positive rate double, which led to a reduced final score – but this was still over 99 and thus the product earns its fifth consecutive VBSpam award.

**Symantec Messaging Gateway 9.5**

- **SC rate:** 99.85%
- **FP rate:** 0.07%
- **Final score:** 99.51
- **Project Honey Pot SC rate:** 99.83%
- **Abusix SC rate:** 99.89%
- **Newsletters FP rate:** 0.0%

It will come as little surprise that Symantec’s Messaging Gateway virtual appliance scans attachments using the Symantec malware scanner. The product can be controlled either via a web GUI or a CLI, allowing for both quick access and a good overview.

The drop in spam catch rate for the virtual appliance was not surprising in the context of this test, but it was nice to see the false positive rate drop too. Therefore, with a
slightly increased final score, the security giant earns its 14th VBSpam award in as many tests.

The Email Laundry
SC rate: 99.73%
FP rate: 0.06%
Final score: 99.45
Project Honey Pot SC rate: 99.66%
Abusix SC rate: 99.84%
SC rate pre-DATA: 98.15%
Newsletters FP rate: 0.0%

The Email Laundry offers its customers multiple MX records spread over multiple locations, which allows for business continuity should one of those locations become temporarily unreachable. Attachments sent through the hosted solution are scanned for malware, though the company declined to reveal which scanner is used.

With a doubled false negative rate, The Email Laundry performed in line with other products, but a decrease in false positive rate meant that its final score was actually improved, and the Irish product achieves its 12th VBSpam award.

Vamsoft ORF
SC rate: 98.62%
FP rate: 0.00%
Final score: 98.62
Project Honey Pot SC rate: 98.71%
Abusix SC rate: 98.49%
Newsletters FP rate: 0.0%

As system administrators can be somewhat religious about their favourite anti-virus software, they will be pleased to learn that ORF has the option for an external anti-malware agent to be included. The Windows solution can be managed using a GUI which, as we have mentioned in the past, gives administrators a lot of control over the product.

Like other products, ORF scored a lower spam catch rate this month, and to see it drop well below 99% was a little disappointing. On the other hand, even though the ham corpus was larger than previously, the product once again blocked no legitimate emails – one of only three full solutions to achieve a zero FP rate this month. The company’s 12th VBSpam award is thus well deserved.

ZEROSPAM
SC rate: 99.90%
FP rate: 0.07%
Final score: 99.56
Project Honey Pot SC rate: 99.94%
Abusix SC rate: 99.83%
SC rate pre-DATA: 96.72%
Newsletters FP rate: 5.9%

ZEROSPAM is the latest product to join the VBSpam tests. The Canadian company offers a hosted anti-spam solution, which can be controlled using a web interface. I played around with this interface a little and was pleased by how easy it was to customize the product to an administrator’s needs and to manage the quarantine. One can switch between English and French at any given time.

The hosted solution gives customers multiple MX records and its hosts are spread over multiple data centres across Europe and Canada; the company explicitly names the
US Patriot Act as a reason to locate all of its data centres outside the US. Attachments sent to the product are scanned using ClamAV.

ZEROSPAM, which uses a statistical classifier, missed only about one in 1,000 spam emails, which gave it one of the highest spam catch rates this month. It blocked only six legitimate emails – impressive for a product for which the whole ham corpus was new – and achieved the fifth highest final score. With these impressive results, ZEROSPAM earns a VBSpam award on its debut appearance.

**Spamhaus ZEN+DBL**

**SC rate:** 96.85%
**FP rate:** 0.00%
**Final score:** 96.85

**Project Honey Pot SC rate:** 97.75%
**Abusix SC rate:** 95.61%
**SC rate pre-DATA:** 95.56%
**Newsletters FP rate:** 0.0%

The drop in spam catch rate seen in all but one product this month was most noticeable in Spamhaus, where it dropped from above 99% to below 97%. While Spamhaus thus fails to win a VBSpam award, it would be wrong to conclude that Spamhaus ‘failed this test’.

Spamhaus Zen+DBL is a ‘partial solution’ that checks the sender’s IP address, their domain and the domains of URLs spotted in the message body against DNS-based blacklists. While in previous tests this has blocked the overwhelming majority of spam, with few to no false positives, the product is generally not used on its own (or intended to be). To understand this, it is important to realise that a small portion of spam is sent from legitimate mail servers and includes links to domains on legitimate web servers – these simply cannot be blocked by domain- or IP-based blacklists without causing false positives.

Whether the drop in Spamhaus’s catch rate was because of an increase in this latter kind of spam, or because the product failed to pick up new IP addresses and domains fast enough is hard to tell. Hopefully the catch rate will pick up again in the next test. With no false positives, even among newsletters, I believe the product continues to be a valuable addition to many an anti-spam solution.
SURBL

SC rate: 56.61%
FP rate: 0.00%
Final score: 56.61
Project Honey Pot SC rate: 46.9%
Abusix SC rate: 70.02%
Newsletters FP rate: 0.0%

Like Spamhaus, the performance of SURBL’s domain-based blacklist deteriorated and, as with Spamhaus, it is hard to tell whether the blacklist has become less effective or whether spammers have become better at using legitimate domains to link to in their emails. Continued reports of compromised websites support the latter theory and it will be interesting to see whether this continues to be reflected in SURBL’s performance. As this usage of compromised websites makes it harder to prevent legitimate domains from appearing in the blacklist, the fact that SURBL once again had no false positives speaks volumes for the hard work put in by its developers.

CONCLUSION

If VBSpam tests were judged by how easy they made my life, this would have been a good test. Lower spam catch rates and more false positives – both in absolute and in relative numbers – give me more to write about. They also make for a more interesting VBSpam graph.

However, the tests don’t exist to make my life easy. And, while it is worth noting that most products still blocked more than 99 out of 100 spam emails, and that no product incorrectly marked more than 1 in 470 legitimate emails as spam2, this is not good. More spam means more time wasted dealing with it, a greater chance of falling for scams, and a greater chance of accidentally deleting legitimate emails.

We hope that participants will be able to demonstrate that this month’s drop in performance is a one-off thing and hope to see catch rates pick back up again in the next test. We will also be adding some more qualitative evaluations of participating products in the next test, giving potential customers an even better picture.

The next VBSpam test will run in April 2012, with the results scheduled for publication in May. Developers interested in submitting products should email martijn.grooten@virusbtn.com.

2 Strictly speaking this is not true. However, we do not count more than four false positives per sender and/or thread – to avoid the results being skewed by one incorrectly blacklisted domain or IP address and to reflect the real-life situation where administrators are usually alerted when legitimate emails continue to be blocked.